

Received April 18, 2019, accepted May 13, 2019, date of publication May 20, 2019, date of current version June 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917604

Generative Adversarial Networks Selection Approach for Extremely Imbalanced Fault Diagnosis of Reciprocating Machinery

DIEGO CABRERA^{1,2}, FERNANDO SANCHO³, JIANYU LONG¹, RENÉ-VINICIO SÁNCHEZ², SHAOHUI ZHANG¹, MARIELA CERRADA², AND CHUAN LI¹, (Senior Member, IEEE)

¹School of Mechanical Engineering, Dongguan University of Technology, Dongguan 523808, China

²GIDTEC, Universidad Politécnica Salesiana, Cuenca 010105, Ecuador

³Department of Computer Science and Artificial Intelligence, Universidad de Sevilla, 41012 Seville, Spain

Corresponding author: Chuan Li (chuanli@dgut.edu.cn)

This work was supported in part by the GIDTEC Research Group of the Universidad Politécnica Salesiana, National Natural Science Foundation of China, under Grant 51605406 and Grant 71801046, in part by the Research Program of the Higher Education of the Guangdong under Grant 2016KZDXM054, and in part by the National Key Research and Development Program of China under Grant 2016YFE0132200.

ABSTRACT At present, countless approaches to fault diagnosis in reciprocating machines have been proposed, all considering that the available machinery dataset is in equal proportions for all conditions. However, when the application is closer to reality, the problem of data imbalance is increasingly evident. In this paper, we propose a method for the creation of diagnoses that consider an extreme imbalance in the available data. Our approach first processes the vibration signals of the machine using a wavelet packet transform-based feature-extraction stage. Then, improved generative models are obtained with a dissimilarity-based model selection to artificially balance the dataset. Finally, a Random Forest classifier is created to address the diagnostic task. This methodology provides a considerable improvement with 99% of data imbalance over other approaches reported in the literature, showing performance similar to that obtained with a balanced set of data.

INDEX TERMS Imbalanced data, GAN, model selection, random Forest, reciprocating machinery.

I. INTRODUCTION

Reciprocating machinery, specifically piston compressors, are among the most used machines in industry [1]. The main task of reciprocating machinery is the generation of compressed air mainly used as a power source for pneumatic systems. However, the already low efficiency of the machinery [2] can be compromised by the presence of mechanical system failures. Additionally, these failures can cause costly damage to the machinery [3], and even physical injuries to humans in its environment [4].

To address this issue, multiple efforts have been employed in the creation of fault diagnosis systems that use signals measured in real time on machinery. For example, in [5], use of Principal Component Analysis to build a compressor diagnoser is proposed. In [6], a method based on decision trees is presented for fault diagnosis. [7] addresses the problem of

fissure detection in valves with time-frequency features and logistic regressor classifiers. Expanding the panorama even further to fault diagnosis in general, some works introduce several approaches based on fuzzy logic [8], [9].

Owing to the large amount of data currently available, intelligent algorithms, especially the methods based on deep learning, have been widely used in the field of intelligent prediction [10] and [11]. For example, in [12], a DBN is used in conjunction with Teager-Kaiser energies for the diagnosis of compressor failures. Although no more works have been reported using deep learning in compressors, in [13] an approach based on deep convolutional autoencoders is presented for the estimation of failure severity in rotating machinery, and [14] proposes the use of deep recurrent models for the task of gear fault diagnosis. Finally, [15] presents a detailed review of works based on machine learning applied to bearing fault diagnosis.

All the previous works have in common a deficiency; that is, they do not consider the real conditions of acquisition

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine.

of signals in industry. When a compressor has a damaged element, it is usually replaced as soon as possible to avoid further damage to the machine. This results in the number of available signals of the different fault modes of the compressor being much lower (2% or even less) compared to those available for healthy conditions. Under this premise, the creation of a diagnostic model should be addressed as a problem of imbalanced data, which, to the best of our knowledge, has not been done before for reciprocating machinery. However, some works have recently reported use of Generative Neural Network (GAN) theory to address this problem for rotating machinery. For example, [16] presents a method for the fault diagnosis of bearings based on a GAN for the increase of data represented in the Fourier spectrum (data are initially balanced in their dataset, but the objective is to increase their quantity to improve the performance of the classifier). One study [17] addresses the fault detection task in electric motors, in which a Deep Convolutional GAN (DCGAN) is used to increase the number of Intrinsic Mode Functions corresponding to vibration signals in the minority fault modes. In [18], the DCGANs are used to create wind turbine fault data from the expert knowledge of the operators of its SCADA system, and the final model only determines whether or not there is a fault in the turbine. Other works include [19] and [20].

The above-mentioned works do not address the problem of extreme imbalance that can occur in multiple tasks of fault diagnosis. At best, they only address 80% of the imbalance problems with decreasing performance with extreme imbalance. In addition, as mentioned before, there are no reported works for reciprocating machinery that address the problem of data imbalance. Therefore, in this work we propose a novel method based on the selection of GAN models for the fault diagnosis task in reciprocating compressors with extreme data imbalance. Our main contributions are (i) a method that combines Wavelet Packet Transform (WPT) with GAN for the creation of Random Forest (RF) classifiers used in the fault diagnosis on compressors with data imbalance of 99%, and (ii) an unsupervised procedure based on similarity of clusters for the selection of the best generative model under conditions of extreme imbalance. To provide robust results, our proposal is compared with seven different methods addressing the same problem (including other GAN-based approaches). Results show that our approach has a near performance (0.986) that can be obtained with the balanced data (0.99), and surpasses the rest of the compared methods.

The rest of this paper is organized as follows. In Section II, we detail the background of techniques used. The proposed methodology is presented in Section III. In Section IV, we introduce the experimental test bed, compare the application of our proposal with other state-of-the-art methods, and discuss the obtained results. Finally, in Section V, we present conclusions and lines of future work that can take advantage of this proposal.

II. BACKGROUND

A. WAVELET PACKAGE DECOMPOSITION

Wavelet analysis [21] is based on the decomposition of a signal into a set of components. Its main difference from Fourier analysis is the use of basic functions, also known as mother wavelets, which are finite-duration functions changing their frequency (as in the classic Fourier analysis) as well as their temporary position. In this way, the transform allows adjustment of the resolution in both time and frequency, obtaining a better resolution in time for high-frequency events and a better resolution in frequency for low-frequency components. Owing to the Heisenberg Uncertainty Principle [22], when crossing a threshold an improvement in resolution over time cannot be obtained without worsening the resolution in frequency, and vice versa.

In this context, the Discrete Wavelet Transform (DWT) used in a discrete signal $v(t)$ is given by

$$c(j, k) = \sum_{t \in \mathbb{Z}} v(t) \psi_{j,k}(t), \quad (1)$$

where $\psi_{j,k}$ is an element of the family basis function ψ given by

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}} \cdot \psi(2^{-j}t - k). \quad (2)$$

The changes in time and frequency are implicitly coupled in Eq. 1, i.e., the translation parameter, τ , and scale parameter, s (reciprocal of frequency), can be computed as

$$\tau = 2^j \quad (3)$$

$$s = 2^j k. \quad (4)$$

Multi-Resolution Analysis (MRA) [23] relates DWT with the recursive application of filters to the signal, decreasing the computational cost of the DWT application. The calculation of the DWT based on the MRA establishes that a signal can be decomposed in its components of high frequency, d (details coefficients), and low frequency, a (approximation coefficients), by applying filters with impulsive response $h(t)$ and $g(t)$, respectively, followed by a sub-sampling operation with factor 2. This is valid as long as the filters are related by $g(L-1-t) = (-1)^t \cdot h(t)$, called quadrature mirror filters, for a signal of duration L . The relationship can be formalized by

$$d(k) = \sum_t v(t) \cdot g(2k - t), \quad (5)$$

$$a(k) = \sum_t v(t) \cdot h(2k - t). \quad (6)$$

The Wavelet Packet Transform (WPT) [24] applies this decomposition process recursively to $a(\cdot)$ and $d(\cdot)$, where every successive application of Eqs. 5 and 6 builds a new decomposition level. This procedure can continue until $a(\cdot)$ and/or $d(\cdot)$ have only one element. However, in practice, the level of decomposition is limited by some optimization criterion or determined by the specific application.

B. GENERATIVE ADVERSARIAL NETWORKS

Generative Adversarial Networks (GANs) [25] are composed of two models, Generator (G) and Discriminator (D), both competing among themselves. The G model receives z input sampled from a known distribution P_z . The result of the application of G to z is an example $x \sim P_g$ where $x = G(z)$.

The D model can receive $x \sim P_r$ input, where P_r is the data distribution, or $x \sim P_g$ input from the G model. The task of D is to recognize whether it is $x \sim P_r$ or $x \sim P_g$, i.e.,

$$D(x) = \begin{cases} 1, & \text{if } x \sim P_r, \\ 0, & \text{if } x \sim P_g \end{cases} \quad (7)$$

The G model must be improved to fool D . The D model improves avoidance of being fooled by G and, at the same time, correctly detects the inputs from P_r . Both improving criteria can be found by the application of maximum-likelihood estimation to a m samples batch of P_r - and P_g -based data entered to a θ -parameterized probabilistic model P_D defined by D model:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \sum_{i=1}^m \log P_D(y_i | x_i; \theta), \quad (8)$$

where $y_i | x_i$ is a known example in the data batch with $y_i = 1$ if $x_i \sim P_r$, and $y_i = 0$ otherwise. As D is a binary classifier, P_D follows a Bernoulli distribution, and Eq. 8 can be reformulated as

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^m \log([D(x_i)]^{y_i} [1 - D(x_i)]^{1-y_i}) \quad (9)$$

$$= \arg\max_{\theta} \sum_{i=1}^m \log[D(x_i)]^{y_i} + \sum_{i=1}^m \log[1 - D(x_i)]^{1-y_i} \quad (10)$$

$$= \arg\max_{\theta} \sum_{i=1}^m y_i \log D(x_i) + \sum_{i=1}^m (1 - y_i) \log [1 - D(x_i)], \quad (11)$$

where D model is parameterized by θ . The left-hand term of Eq. 11 refers to the ability of D to recognize $x_i \sim P_r$ and the other term its ability to reject $G(z_i) = x_i \sim P_g$. Taking this into account, Eq. 11 can be reformulated as the optimization of the following two expectations:

$$\hat{\theta} = \arg\max_{\theta} \mathbb{E}_{x \sim P_r} \log D(x) + \mathbb{E}_{z \sim P_z} \log [1 - D(G(z))], \quad (12)$$

where the first summand improves the model's ability to detect real inputs and the second summand improves the model's ability to recognize inputs generated by G .

If G is parameterized by ϕ , the previous criterion must be minimized to find the appropriated parameters that improve this model. However, the first expectation is independent of G and can be removed. Finally, the optimization criteria for G can be formulated as

$$\hat{\phi} = \arg\min_{\phi} \mathbb{E}_{z \sim P_z} \log [1 - D(G(z))], \quad (13)$$

showing that the generator's improvement does not disturb the ability of the discriminator to recognize $x \sim P_r$, but it is related to the ability to detect $x \sim P_g$.

Other improvements have been proposed to avoid model-collapse issues in GAN training. One of the most popular is the use of Wasserstein loss functions [26] for $\hat{\theta}$ and $\hat{\phi}$ optimization. In this case, the discriminator is replaced by the critic function f , which must have 1-Lipschitz continuity property and be parameterized by θ . The optimization criteria for f and G are then

$$\hat{\theta} = \arg\max_{\theta} \mathbb{E}_{x \sim P_r} f(x) - \mathbb{E}_{z \sim P_z} f(G(z)), \quad (14)$$

$$\hat{\phi} = \arg\min_{\phi} -\mathbb{E}_{z \sim P_z} f(G(z)). \quad (15)$$

C. RANDOM FOREST

Random Forest (RF) [27] is a machine-learning model that has the advantages of classification and regression trees [28] by grouping a large number of them. Each tree of the RF model works independently of the others and, finally, some technique of aggregation of results is applied to obtain a final conclusion.

RF is an example of ensemble machine-learning theory. In this theory, an ensemble model, h , is composed of K submodels, h_j , called weak learners:

$$h = \{h_1(x), h_2(x), \dots, h_j(x), \dots, h_K(x)\}. \quad (16)$$

Weak learners give an independent response to a specific instance that must be classified/regressed. This means that the parameters associated with each learner also have independence from the others. For example, if weak learners belong to the same family of models, then we can write $h_j(x) = h(x|\theta_j)$, where θ_j represents the parameters associated with the j th learner, i.e., architecture, input samples, input features, etc.

The outputs of the weak learners are then combined in some way and the aggregated result is given as the output of the ensemble. For example, in a classification task when using the majority voted aggregator the empirical probability of a specific c_i class is obtained from

$$\hat{P}(c_i|x) = \frac{1}{K} \sum_{j=1}^K I(h(x|\theta_j) = c_i), \quad (17)$$

where $I(\cdot)$ is the indicator function that returns 1 if its argument is true, and 0 otherwise. Finally, the class chosen as result of h is the one with the highest empirical probability distribution.

In RF, two techniques are commonly used to increase the diversity of the generated weak learners:

- Bagging [29]: Denotes bootstrap [30] aggregating. It consists of building the multiple different decision-tree weak learners from the original training dataset by repeatedly using multiple bootstrapped subsets of the data.

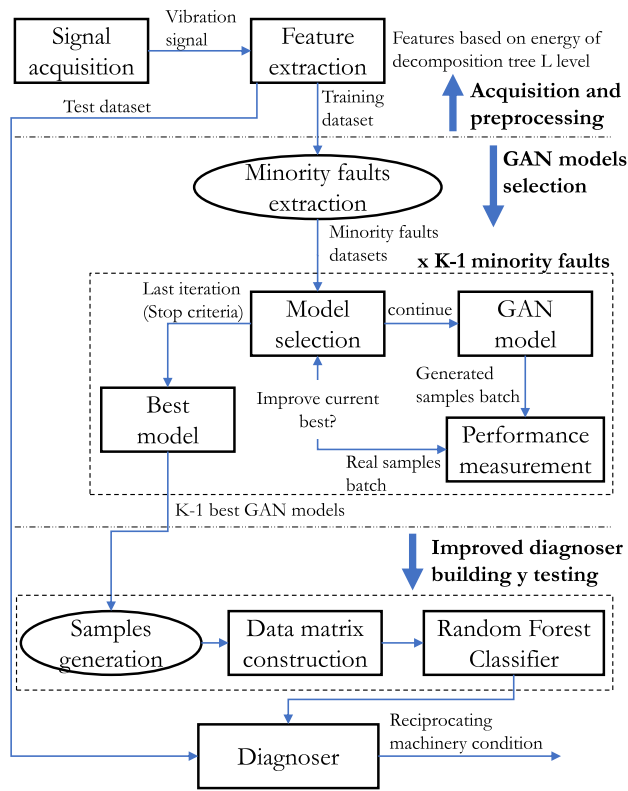


FIGURE 1. Overview of proposed approach.

- **Random Feature Selection:** From all the possible features of the original training dataset, we randomly select a subset of to build each weak learner.

These techniques, in addition to enriching the model with different learners, also decrease its computational complexity.

III. GAN-BASED DATA BALANCE FOR FAULT DIAGNOSIS

In this section we present a novel GAN-based approach for solving the problem of highly imbalanced datasets commonly obtained in real-world fault diagnosis applications of reciprocating machinery. The samples are composed of features extracted from the approximations and details coefficients of the WPT. Then, the GAN models assess the data distribution for every minority faulty mode to synthetically increase its size. As is well known, the optimization process of GAN models suffers from two main issues: (i) instability due to competition between generator and discriminator models, and (ii) a subjective criterion in the evaluation. To deal with these problems, in this work we introduce an online unsupervised model-selection stage in the training process of the GAN based on a statistical similarity index between the cluster of real samples and the cluster of generated samples. Finally, a RF classifier is built from the balanced dataset and applied to diagnose the fault. Figure 1 summarizes this proposal.

A. SIGNAL ACQUISITION AND PRE-PROCESSING

Suppose that a vibration signals set obtained under K different healthy states of the machinery is available, and let $\{P_1, \dots, P_K\}$ be the family of sets of different machinery states, where P_1 contains samples under the normal condition (where all components are healthy) and P_i contains samples with the same faulty state (for $i = 2, \dots, K$).

Usually, P_1 is the larger set due to the ease of capturing signals under this condition, and we will assume that the relative sizes of the other sets with respect to P_1 are very small ($|P_i| \leq c \cdot |P_1|$, $i = 2, \dots, K$ with $c \ll 1$) to test our methodology in a case of real imbalance (for example, in Section IV we present cases in which $c = 0.01$).

Let $s(t)_{t \in [0, T]}$ be a discrete vibration signal (a time-series), as shown in Fig. 2. Each element in $s(t)$ is temporally dependent on its predecessors, and consequently their elements are highly correlated, but not suitable as inputs for shallow machine-learning models like RF.

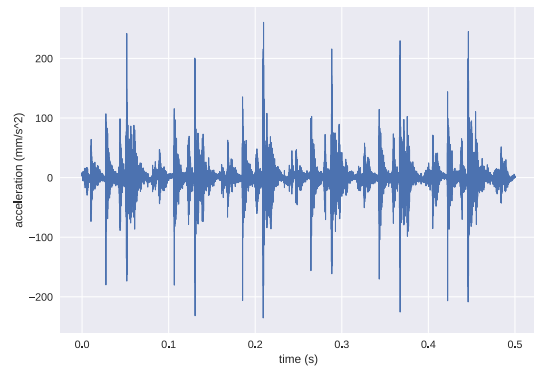


FIGURE 2. Example of vibration signal from reciprocating machinery.

To deal with this issue, we propose the use of WPT as decomposition method for enhanced information of raw signal. This approach has been previously successfully tested in other fault diagnosis studies [31]. First, the decomposition of the vibration signal is carried out through the recursive application of wavelet quadrature filters. This procedure continues until a L layer of decomposition is reached, creating the so-called decomposition tree, where a 2^L approximation and detailed coefficients signals are obtained. Finally, the coefficients signal energy is computed for each node in the last layer, obtaining a features vector of 2^L components. An example of this procedure is presented in Fig 3.

For every vibration signal s , we denote x_s as the feature vector extracted by the application of the previous procedure. Then, for every $i \in \{1, \dots, K\}$ we denote

$$X_i = \{x_s : s \in P_i\}. \quad (18)$$

B. GAN MODEL SELECTION

As studied in [32], the training process of GAN models suffer from instability and a collapse mode. The last issue can

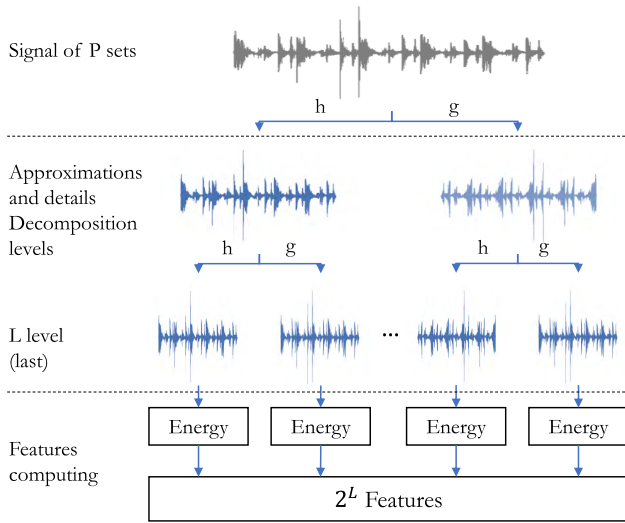


FIGURE 3. Wavelet packet decomposition tree-based feature extraction for vibration signals.

be addressed by Wasserstein GAN models as presented in this work. However, the instability of the training process continues to be an open issue for this type of model. Compounding this is the model's inability to objectively evaluate the performance of the generator in tasks other than image generation. As will be seen in Section IV, the scope of the Nash equilibrium is not an indicator of good performance. Owing to these problems, the generator model can decrease its performance in the task after a few iterations without it being noticed by the training algorithm.

To address the aforementioned issue, we propose introducing an unsupervised mechanism of model selection within the common GAN training algorithm. The proposal is guided by the dissimilarity of the clusters formed by two data batches, one of real data, B_r , and the other of generated data, B_g .

As a first approximation, in this work we use the next basic definition, but it would be interesting to analyze how other measures affect the performance of the methodology:

$$\text{diss}(B_r, B_g) = \|C_r - C_g\|_2 + \|S_r - S_g\|_2, \quad (19)$$

where C_r (respectively, C_g) is the centroid of the real (respectively, generated) data cluster, and S_r (respectively, S_g) is the real (respectively, generated) data sparsity (specifically, standard deviation).

We follow a greedy mechanism. In each training iteration, the selection mechanism verifies whether there is an improvement in the generator model with respect to the dissimilarity measure. A lower measure indicates a better model, as shown in Fig 4. In this case, the model is stored, replacing the previous one best model. This process continues for a number of predetermined iterations, and it is repeated for each minority fault state. Algorithm 1 summarizes this procedure. As a result of its application, $K - 1$ generative models have been saved for their use at the next stage.

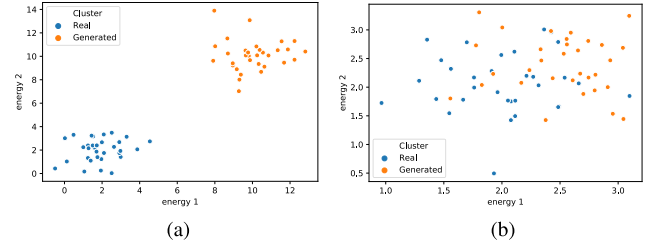


FIGURE 4. Examples of dissimilarity indexes between real and generated clusters. (a) Clusters with dissimilarity of 11.31. (b) Clusters with dissimilarity of 0.707.

Algorithm 1 GAN Models Building and Selection

Data: $\{P_1, \dots, P_K\}$

Result: Best GAN models collection for wavelet-based examples generation

```

1 foreach  $i = 2 \dots K$  do
2    $f, G_i(\theta, \phi)$ : parameterized models;
3   Init  $\theta$  and  $\phi$ ;
4    $best = \infty$ ;
5   foreach iteration in training do
6      $B_r$  = Random batch from  $X_i$ ;
7      $B_g$  = Random batch from  $G_i$ ;
8     Train  $\theta$  through  $B_r$  and  $B_g$  (Eq. 14);
9      $B_r$  = Random batch from  $X_i$ ;
10     $B_g$  = Random batch from  $G_i$ ;
11    Train  $\phi$  through  $B_r$  and  $B_g$  (Eq. 15);
12     $B'_g$  = Random batch from  $G_i$ ;
13    Compute  $d = \text{diss}(B_r, B'_g)$  (Eq. 19);
14    if  $d < best$  then
15      Store  $G_i$  as generator for  $P_i$ ;
16       $best = d$ ;

```

Selected architecture for the generator and the discriminator are both the multilayer perceptron neural network, because of the power of this architecture to naturally represent time-independent characteristics (generator) as well as its ability to classify them (discriminator/critic) (see Figure 5).

C. IMPROVED DIAGNOSER BUILDING AND TESTING

After obtaining the collection of generative models from the previous algorithm, we use them to generate new synthetic samples and complete several fault state datasets, i.e., for every $2 \leq i \leq K$, we use the G_i model to generate as many synthetic states as we need to obtain a new dataset X'_i verifying $X_i \subseteq X'_i$ and $|X'_i| = |X_1|$. In this way, the new global dataset $X' = \{X_1, X'_2, \dots, X'_K\}$ is balanced.

Finally, an RF classification model is built with the balanced X' training set. As we demonstrate in the next section, this classifier performs better than the one that can be built with the imbalanced dataset. The procedure for creating the weak learners follows the standard procedure.

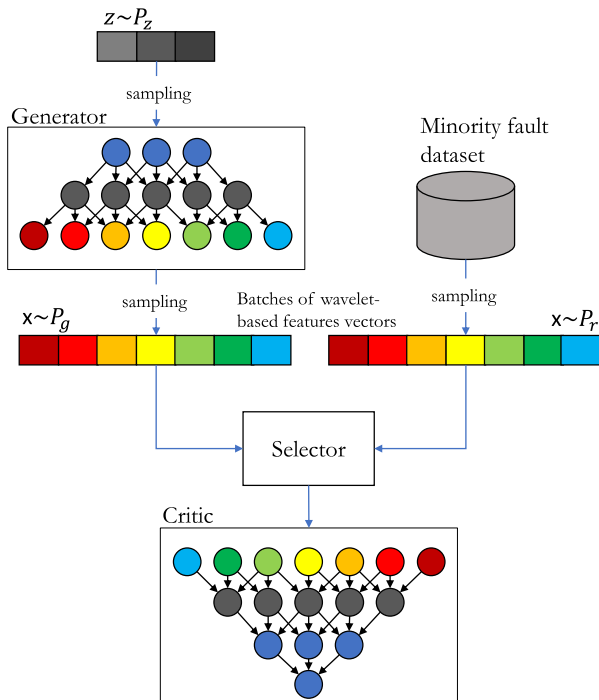


FIGURE 5. Architecture of GAN models.

Once the classifier has been obtained, the instances of the test set can be evaluated, and, additionally, new time series can be classified online if they first go through the feature-extraction phase and then entered into the classifier. Note that in this stage the generator models are not required, as they were used only to balance the training set.

IV. EXPERIMENTS

A. TEST BED AND SIGNALS DATASET

The experiments were performed in the reciprocating machinery test bed shown in Fig. 6 and implemented in the laboratory of GIDTEC Research Group of the Universidad Politécnica Salesiana (Cuenca, Ecuador). Its main part is the EBG250 two-stage reciprocating compressor 5 HP. A three-phase motor is coupled to a belt-driven transmission system to provide force to the compression camera. The motor rotation speed is 57.7 Hz, which results in a crankshaft rotation frequency of 12.8 Hz.

The measurement of the vibration signal was performed with a 100mV/g sensitivity IMI SENSORS 603C01 accelerometer sensor. It was vertically located over the compression camera. Then, the analog signal was transferred by shielded cable to a National Instrument NI9234 data-acquisition card, where the analog-to-digital conversion is performed at a sampling rate of 50 kS/s. This card was mounted on a NI9188 chassis of the same brand for the purpose of transferring the data to a management and storage system programmed in a computer.

The vibration signals were acquired under healthy conditions and four different failure conditions in the compressor

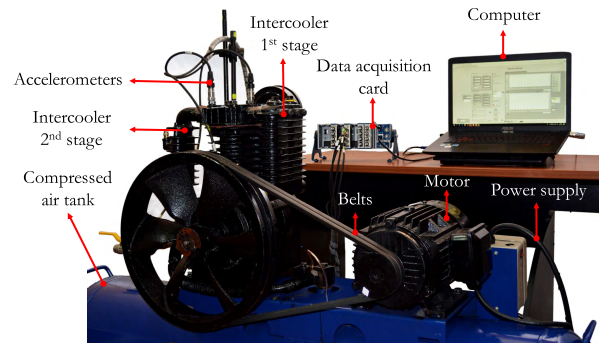


FIGURE 6. Test bed of two-stage reciprocating compressor.

TABLE 1. Distribution of signals for fault configurations.

Code	Description	# Training set	# Test set
P_1	Healthy condition	24510	5588
P_2	Discharge valve seat wear	245	5588
P_3	Intake valve corrosion	245	5588
P_4	Discharge valve crack	245	5588
P_5	Intake valve spring crack	245	5588

valve subsystems that were identified as the most critical. The failure modes selected were valve seat wear 1.44 mm, valve plate corrosion with diameter of 2.5 mm, valve crack with diameter of 1.6 mm and spring break, as shown in Fig. 7.

For each condition, 100 and 50 vibration signals of 10s duration were acquired at a pressure between 2.9–3.0 bar for the train and test stages, respectively. Then, from each signal, sub-signals with a random starting point and of 1s duration were extracted. For the extremely imbalanced case, only one of the former signals was used for the fault modes. Table 1 summarizes the fault configuration and the number of signals obtained for each dataset (as it can be seen, the set of signals in fault condition represents only 1% of the set of signals in normal condition for the training stage). However, the set used for the model test is balanced to obtain a fair evaluation.

B. COMPARATIVE METHODS

In the following, the feature extraction of signals through WPT is configured with Daubechies 7 and Symlet 3 wavelets as the best wavelet families for fault diagnosis according to [33], at six levels of decomposition, obtaining 64 energy values in the last level per wavelet. The resulting vectors are concatenated, obtaining 128 features.

To make the comparisons, we first used the RF model created with the balanced dataset (RF-B), i.e., 24,510 examples for each fault condition, and the RF model created with the unbalanced data set (RF-I). In this sense, RF-B and RF-I are the baseline models for comparisons, respectively our upper bound and lower bound of accuracy.

Following [34], the first methods addressing the problem of imbalance applied to our case study are oversampling and undersampling by random replication of examples. In the



FIGURE 7. Healthy condition and fault modes in components. (a) Healthy condition of valve components. (b) Valve seat wear. (c) Valve plate corrosion. (d) Valve plate crack. (e) Spring break.

first method, a duplication of randomly selected elements is made in the undersampled failure sets until reaching the balance in the entire dataset. In the second method, we follow the inverse process, and examples of the majority set are randomly selected to be eliminated until the balance is reached. With the new training sets, new RF models are created, obtaining RF-O (for the oversampling method) and RF-U (for the undersampling method).

The next balancing methods to be compared are SMOTE [35] (Synthetic Minority Over-sampling Technique) and its evolution, ADASYN [36] (Adaptive Synthetic sampling). SMOTE creates vectors between each example of the minority datasets and its k -nearest neighbors (in the same dataset). Then, each vector is multiplied by a random constant between 0 and 1, and the resulting point is the new example added to the dataset. This process is repeated until the balance of the minority fault modes is reached. The improvement introduced by ADASYN to SMOTE lies in adding noise to the new examples to avoid linear correlation with the parent instances. Both are oversampling methods. As before, after balancing the training dataset, new RF classifiers are built, RF-S for SMOTE and RF-A for ADASYN.

Finally, we create three more classifiers to evaluate the incidence of GAN model selection:

- RF-BGAN1: trained from 24,510 examples synthetically generated for each minority failure mode, i.e., the entire dataset is synthetic (pure synthetic with best model selection).

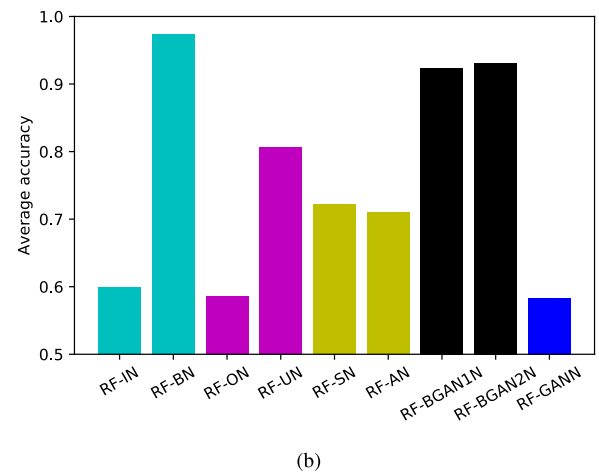
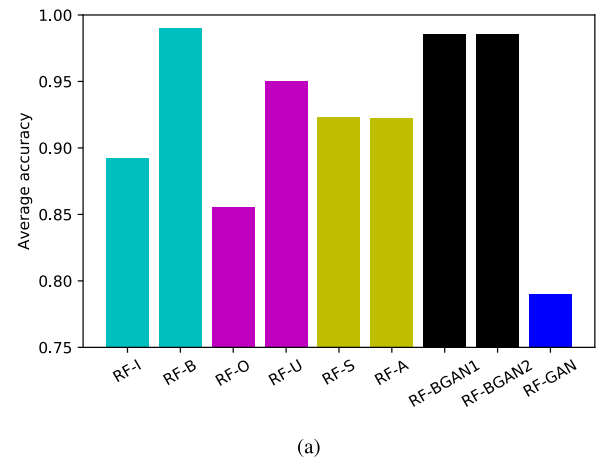


FIGURE 8. Average accuracy for unnormalized and normalized data input for each classifier. (a) Average accuracy for unnormalized input. (b) Average accuracy for normalized input.

- RF-BGAN2: trained from 24,265 examples synthetically generated for each minority failure mode, combined with the 245 original examples of the former dataset (mixed synthetic/original with best model selection).
- RF-GAN: trained with 24,526 examples from the resulting GAN at last iteration (step 8000) as presented in [19], combined with the 245 original examples of the former dataset (mixed synthetic/original with no model selection).

In all three models the neural network on the generator model has a linear output layer to generate features in an extended range surpassing the $[-1, 1]$ values. In addition, the critic model has the linear output layer as required by the Wasserstein loss. For RF-BGAN1 and RF-BGAN2, the best generator/critic hyper-parameters for multilayer perceptrons, such as number of layers, number of units, batch size, and z length vector, are obtained through the application of grid-search methodology [37]. An Adam optimization algorithm [38] is applied to optimize the weights sets of all neural networks, and learning rates are fixed to 1.0×10^{-4} and 1.0×10^{-5} .

TABLE 2. Precision of classifiers for each machinery condition. Average over 20 repetitions together with confidence interval is presented.

Classifier	P_1	P_2	P_3	P_4	P_5
RF-I	0.634 \pm 0.001	0.988 \pm 0.001	0.994 \pm 0.0	0.999 \pm 0.0	1.0 \pm 0.0
RF-B	0.996 \pm 0.0	0.976 \pm 0.0	0.984 \pm 0.0	1.0 \pm 0.0	0.998 \pm 0.0
RF-O	0.56 \pm 0.002	0.998 \pm 0.0	0.993 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0
RF-U	0.931 \pm 0.001	0.942 \pm 0.001	0.925 \pm 0.001	0.941 \pm 0.0	0.955 \pm 0.0
RF-S	0.712 \pm 0.003	0.997 \pm 0.0	0.987 \pm 0.001	1.0 \pm 0.0	1.0 \pm 0.0
RF-A	0.712 \pm 0.002	0.997 \pm 0.0	0.986 \pm 0.001	1.0 \pm 0.0	1.0 \pm 0.0
RF-BGAN1	0.984 \pm 0.0	0.97 \pm 0.001	0.977 \pm 0.0	1.0 \pm 0.0	0.999 \pm 0.0
RF-BGAN2	0.987 \pm 0.0	0.97 \pm 0.001	0.977 \pm 0.0	1.0 \pm 0.0	0.999 \pm 0.0
RF-GAN	0.305 \pm 0.1	0.999 \pm 0.001	1.0 \pm 0.0	1.0 \pm 0.0	0.999 \pm 0.0

TABLE 3. Recall of classifiers for each machinery condition. Average over 20 repetitions together with confidence interval is presented.

Classifier	P_1	P_2	P_3	P_4	P_5
RF-I	1.0 \pm 0.0	0.632 \pm 0.003	0.863 \pm 0.002	0.994 \pm 0.0	0.986 \pm 0.0
RF-B	0.972 \pm 0.0	0.981 \pm 0.0	0.995 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0
RF-O	1.0 \pm 0.0	0.694 \pm 0.003	0.692 \pm 0.004	0.989 \pm 0.0	0.984 \pm 0.0
RF-U	0.921 \pm 0.001	0.952 \pm 0.002	0.931 \pm 0.001	0.965 \pm 0.0	0.911 \pm 0.0
RF-S	1.0 \pm 0.0	0.844 \pm 0.001	0.833 \pm 0.003	0.993 \pm 0.0	0.989 \pm 0.0
RF-A	1.0 \pm 0.0	0.844 \pm 0.001	0.833 \pm 0.002	0.993 \pm 0.0	0.989 \pm 0.0
RF-BGAN1	0.979 \pm 0.001	0.964 \pm 0.001	0.986 \pm 0.0	0.997 \pm 0.0	0.999 \pm 0.0
RF-BGAN2	0.979 \pm 0.001	0.965 \pm 0.0	0.987 \pm 0.0	0.998 \pm 0.0	1.0 \pm 0.0
RF-GAN	1.0 \pm 0.0	0.632 \pm 0.003	0.743 \pm 0.002	0.814 \pm 0.0	0.776 \pm 0.0

for generator and critic, respectively. The number of training iterations is fixed at 8000 steps.

Additionally, the normalization incidence (mean 0, standard deviation 1) of income data based on wavelet features is evaluated for all previously described methods, including our proposal. In this sense, for every method we build two classifiers, one with no normalized data (for example, RF-B) and a similar one that uses normalized data (for example, RF-BN).

Since the execution of the previous methods depends on randomness, to avoid bias of the results 20 models are created under the same configuration for each method. The hyper-parameters of these models, such as *number of trees* and *number of random selected features*, are optimized by exhaustive search. The percentage of bagging is maintained at 70%, and the stop criterion for tree building is *maximum depth*. Hence, each classifier is evaluated with the balanced test set, and the results are averaged to obtain a final conclusion.

C. COMPARATIVE RESULTS

Figure 8 compares the methods for the imbalance of data proposed in the literature together with our method: Fig. 8 (a) shows the results for models using the unnormalized inputs, and Fig. 8 (b) those for models using the normalized inputs.

Results of Fig. 8 (a) show that RF-BGAN1 (0.985) and RF-BGAN2 (0.986) are the models closest to the upper bound, with only a difference of 0.005 and 0.004, respectively. However, the minimum difference (approximately 0.001) between

the accuracy of the two models is evidence that all the discriminant information existing in the data of the minority failures is contained in the data generated by the GAN. Therefore, the improvement in accuracy achieved with the union of the real and generated data is negligible.

The next-best model is RF-U, surpassing any more elaborate technique such as SMOTE (RF-S), ADASYN (RF-A), and even GAN without model selection (RF-GAN), which is in last place (worse than the lower bound obtained with RF-I). This shows the robustness of RF with a limited number of examples when providing the appropriate features. However, RF-O shows that data balancing by simply copying of examples only worsens the final performance of the model. Figure 8 (b) illustrates the same ranking in the performance of methods, but all the methods behave worse than the corresponding unnormalized one. Thus, the typical rule of data normalization is not applicable to this concrete case.

Poor RF-GAN performance can be explained by analyzing the training curves of the model, as in Fig. 9, in which the curves displayed belong to the GAN model for P_3 . We can observe three interesting areas for optimization with minimum values in the proposed dissimilarity index in Fig. 9 (a). The first and second zones are between iterations [100, 500] and [2700, 3000], respectively, and the third between iterations [4000, 5000]. It is clear that the first and second zones are unwanted local minimums. However, the third zone, despite containing the best minimum (0.92 at iteration 4596), is highly variable, showing a sharp landscape close to the optimum in the parameter space of the generator. Thus, selecting the generator model in a uninformed way, only taking

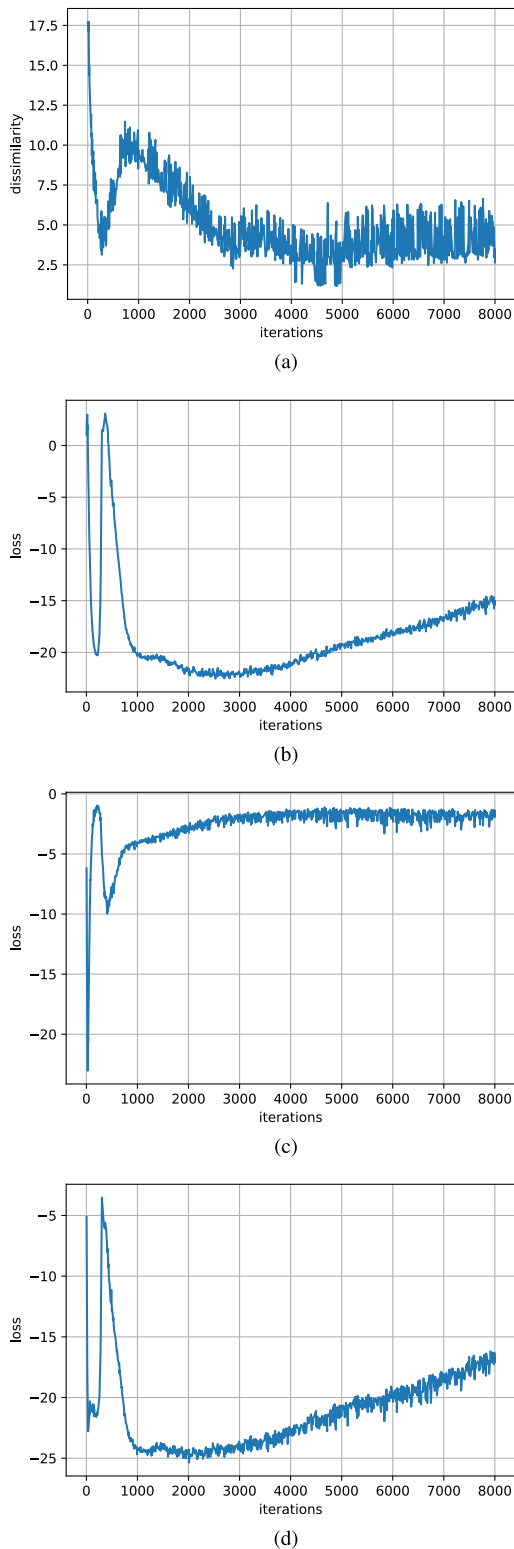


FIGURE 9. Training curves of GAN model for P_3 samples generation. (a) Dissimilarity index. (b) Generator loss. (c) Critic loss. (d) Generator plus critic loss.

into account the number of iterations, produces sub-optimal results. In addition, the curves of the loss function of the generator, critic, and their sum (Fig. 9 (b), 9 (c), and 9 (d),

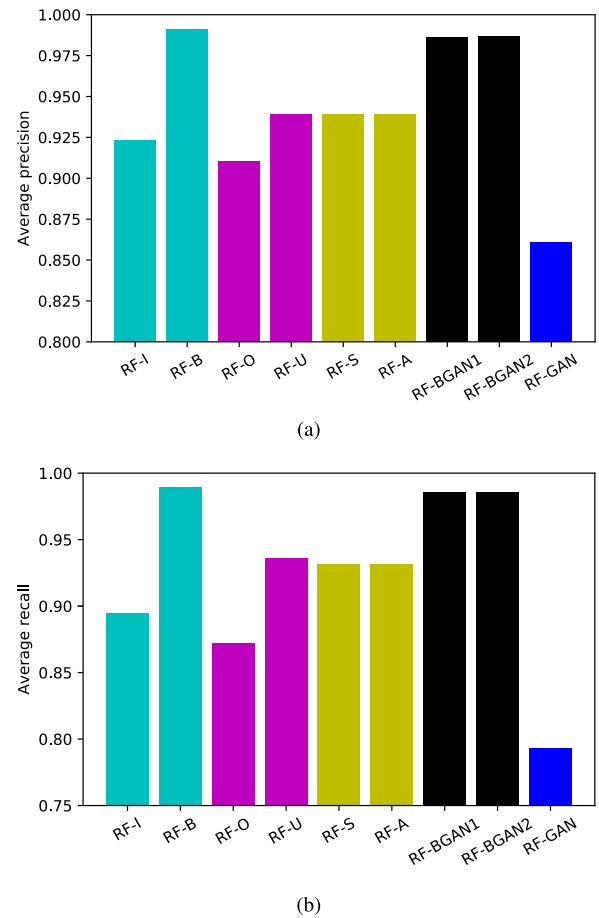


FIGURE 10. Average precision and recall for compared methods. (a) Average precision. (b) Average recall.

respectively) indicate that they cannot be considered for the selection of the best model because at its optimal values there is no information to make a decision about the best generator. This justifies the inclusion of the selection process based on the dissimilarity measure.

For a more detailed comparison, Tables 2 and 3 show the precision and recall results with their confidence intervals in each fault mode. Table 2 shows that the decrease in accuracy using RF-I, RF-O, RF-S, RF-A, and RF-GAN occurs mainly in P_1 . Therefore, a large number of examples of fault modes are confused by the classifier with the healthy condition, which is not acceptable in an early detection task. This confusion could be due to the fact that failure modes P_2 to P_4 are not severe and are considered initial stages of fault mode. Table 3 illustrates that for the same models a good ability to detect the healthy condition is obtained (the examples of this condition are not confused with fault modes), but their performance decreases radically when trying to detect fault modes. Taking into account these two points, we can conclude that the decrease in the recall for fault modes (Table 3) is due to examples classified in P_1 (Table 2), evidencing the deficiency of the examples of the minority groups with which

the classifiers were created. This fact is not visible for RF-U, RF-BGAN1, and RF-BGAN2. In the first model, this is due to the fact that no new examples are added, and in the other two models it is due to the quality of the generated examples. Finally, Fig. 10 illustrates the average precision and recall for the compared approaches.

V. CONCLUSIONS

In this work, a new method of selecting GAN models for the fault diagnosis of reciprocating machinery with highly imbalanced data is proposed. After the signal-acquisition stage, feature extraction using the WPD tree is presented. Next, the creation of GANs with a novel approach to the selection of generative models is introduced as the main contribution. The generators obtained were used to increase the size of the minority fault data and balance the dataset. Then, an RF classifier was created to address the diagnostic task and provide an objective evaluator.

The proposed approach has been evaluated in the diagnosis of five different conditions in the valve sub-system of a reciprocating compressor, and exhaustively compared to several other methods reported in the literature. The results obtained allow us to make the following observations:

- Our proposal improves the classical approaches of undersampling and oversampling, and produces diagnosers almost as powerful as those produced with the balanced dataset.
- For the diagnosis task, the inclusion of model selection in the creation of generators is mandatory for a substantial improvement of the RF diagnoser by the sharpness of the parameter space near the zone of the global optimum. Reaching the last iteration does not guarantee the best model.
- Normalization of the input dataset does not apply to the problem addressed. This may be due to the intrinsic properties of the resulting wavelet packet features, such as the non-Gaussianity of the data.

ACKNOWLEDGMENT

The experimental work was developed at the GIDTEC Research Group Lab of the Universidad Politécnica Salesiana, Ecuador.

REFERENCES

- [1] M. Yang, "Air compressor efficiency in a vietnamese enterprise," *Energy Policy*, vol. 37, no. 6, pp. 2327–2337, Jun. 2009. doi: 10.1016/j.enpol.2009.02.019.
- [2] A. P. Senniappan, "Baselining a compressed air system: An expert systems approach," M.S. thesis, College Eng. Mineral Resour., West Virginia Univ., Morgantown, WV, USA, 2004.
- [3] V. T. Tran, F. AlThobiani, and A. Ball, "An approach to fault compressor valves using Teager-Kaiser energy operator and diagnosis of reciprocating deep belief networks," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4113–4122, Jul. 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417413010014>
- [4] C. J. Guerra, "Condition monitoring of reciprocating compressor valves using analytical and data-driven methodologies," Ph.D. dissertation, Dept. Mech. Eng., Rochester Inst. Technol., Rochester, NY, USA, 2013.
- [5] P. Potočník and E. Govekar, "Semi-supervised vibration-based classification and condition monitoring of compressors," *Mech. Syst. Signal Process.*, vol. 93, pp. 51–65, Sep. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S08832701730047X>
- [6] S. Aravinth, K. R. Kanna, and V. Sugumaran, "Air compressor fault diagnosis through vibration signals using statistical features and J48 algorithms," *Indian J. Sci. Technol.*, vol. 9, no. 47, pp. 6–10, 2016.
- [7] K. Pichler, E. Lughofer, M. Pichler, T. Buchegger, E. P. Klement, and M. Huschenbett, "Fault detection in reciprocating compressor valves under varying load conditions," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 104–119, Mar. 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S088327015004008>
- [8] C. Li, J. L. V. de Oliveira, M. C. Lozada, D. Cabrera, V. Sanchez, and G. Zurita, "A systematic review of fuzzy formalisms for bearing fault diagnosis," *IEEE Trans. Fuzzy Syst.*, to be published. doi: 10.1109/TFUZZ.2018.2878200.
- [9] C. Li et al., "A comparison of fuzzy clustering algorithms for bearing fault diagnosis," *J. Intell., Fuzzy Syst.*, vol. 34, no. 6, pp. 3565–3580, Jun. 2018. doi: 10.3233/JIFS-169534.
- [10] Y. Bai, Y. Li, B. Zeng, C. Li, and J. Zhang, "Hourly PM_{2.5} concentration forecast using stacked autoencoder model with emphasis on seasonality," *J. Cleaner Prod.*, vol. 224, pp. 739–750, Jul. 2019.
- [11] L. Jianyu, S. Zhenzhong, P. M. Pardalos, H. Ying, Z. Shaohui, and L. Chuan, "A hybrid multi-objective genetic local search algorithm for the prize-collecting vehicle routing problem," *Inf. Sci.*, vol. 478, pp. 40–61, Apr. 2019.
- [12] V. T. Tran, F. AlThobiani, T. Tinga, A. Ball, and G. Niu, "Single and combined fault diagnosis of reciprocating compressor valves using a hybrid deep belief network," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 232, no. 20, p. 0954406217740929, Nov. 2017. doi: 10.1177/0954406217740929.
- [13] D. Cabrera, F. Sancho, M. Cerrada, R.-V. Sanchez, and F. Tobar, "Echo state network and variational autoencoder for efficient one-class learning on dynamical systems," *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3799–3809, Jun. 2018. doi: 10.3233/JIFS-169552.
- [14] R. Yang, M. Huang, Q. Lu, and M. Zhong, "Rotating machinery fault diagnosis using long-short-term memory recurrent neural network," *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 228–232, 2018. doi: 10.1016/j.ifacol.2018.09.582.
- [15] M. Cerrada et al., "A review on data-driven fault severity assessment in rolling bearings," *Mech. Syst. Signal Process.*, vol. 99, pp. 169–196, Jan. 2018.
- [16] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213–222, Oct. 2018.
- [17] Y. O. Lee, J. Jo, and J. Hwang, "Application of deep neural network and generative adversarial network to industrial maintenance: A case study of induction motor fault detection," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 3248–3253.
- [18] J. Liu, F. Qu, X. Hong, and H. Zhang, "A small-sample wind turbine fault detection method with synthetic fault data using generative adversarial nets," *IEEE Trans. Ind. Informat.*, to be published.
- [19] W. Mao, Y. Liu, L. Ding, and Y. Li, "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study," *IEEE Access*, vol. 7, pp. 9515–9530, 2019.
- [20] Y. Xie and T. Zhang, "Imbalanced learning for fault diagnosis problem of rotating machinery based on generative adversarial networks," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 6017–6022.
- [21] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989. doi: 10.1109/34.192463.
- [22] W. Heisenberg, "Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik," *Zeitschrift Physik*, vol. 43, nos. 3–4, pp. 172–198, Mar. 1927. doi: 10.1007/BF01397280.
- [23] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.
- [24] M. Y. Gokhale and D. K. Khunduja, "Time domain signal analysis using wavelet packet decomposition approach," *Int. J. Commun., Netw. Syst. Sci.*, vol. 3, no. 3, pp. 321–329, 2010. doi: 10.4236/ijcns.2010.33041.
- [25] J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661v1*. [Online]. Available: <https://arxiv.org/abs/1406.2661>

- [26] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: <https://arxiv.org/abs/1701.07875>
- [27] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001. doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [28] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Monterey, CA, USA: Wadsworth and Brooks, 1984.
- [29] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, vol. 1. Berlin, Germany: Springer, 2001.
- [30] B. Efron and R. J. Tibshirani, *An Introduction to Bootstrap*. Boca Raton, FL, USA: CRC Press, 1994.
- [31] D. Cabrera et al., "Fault diagnosis of spur gearbox based on random forest and wavelet packet decomposition," *Frontiers Mech. Eng.*, vol. 10, no. 3, pp. 277–286, Sep. 2015.
- [32] N. Kodali, J. Abernethy, J. Hays, and Z. Kira, "On convergence and stability of GANs," 2017, *arXiv:1705.07215*. [Online]. Available: <https://arxiv.org/abs/1705.07215>
- [33] M. Cerrada, R. Sánchez, D. Cabrera, G. Zurita, and C. Li, "Multi-stage feature selection by using genetic algorithms for fault diagnosis in gearboxes based on vibration signal," *Sensors*, vol. 15, no. 9, pp. 23 903–23 926, Sep. 2015.
- [34] J. Hernandez, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "An empirical study of oversampling and undersampling for instance selection methods on imbalance datasets," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Berlin, Germany: Springer, 2013, pp. 262–269.
- [35] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
- [36] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1322–1328.
- [37] D. Montgomery, *Design and Analysis of Experiments: Student Solutions Manual*. Hoboken, NJ, USA: Wiley, 2008. [Online]. Available: <http://books.google.de/books?id=kMMJAm5bD34C>
- [38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: <https://arxiv.org/abs/1412.6980>



JIANYU LONG received the B.E. degree and the Ph.D. degree in metallurgical engineering from Chongqing University, Chongqing, China, in 2012 and 2017, respectively. He is currently a Teacher with the College of Mechanical Engineering, Dongguan University of Technology, Dongguan, China. From 2015 to 2016, he was a Visiting Scholar with the Center for Applied Optimization, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA. His research interests include production scheduling, prognostics, and system health management.



RENÉ-VINICIO SÁNCHEZ received the B.S. degree in mechanical engineering from the Universidad Politécnica Salesiana (UPS), Ecuador, in 2004, the Ph.D. degree in industrial technologies research from the UNED, Spain, in 2017. He is currently a Professor with the Department of Mechanical Engineering, UPS. His research interests include machinery health maintenance, pneumatic and hydraulic systems, artificial intelligence, and engineering education.



SHAOHUI ZHANG received the Ph.D. degree in mechatronic engineering from the South China University of Technology, Guangzhou, China, in 2014. He is currently a Teacher with the Mechatronic Engineering Department, Dongguan University of Technology, Dongguan, China. His research interests include mechanical vibration, signal processing, mechanical equipment condition monitoring, and fault diagnosis.



DIEGO CABRERA received the B.Sc. degree in electronic engineering from the Universidad Politécnica Salesiana, Ecuador, in 2012, and the M.Sc. degree in logic, computation, and artificial intelligence and the Ph.D. degree in computer science from Seville University, Spain, in 2014 and 2018, respectively. He is currently a Postdoctoral Fellow with the Dongguan University of Technology, China. He has been a Professor with the Universidad Politécnica Salesiana, since 2014. His research interests include intelligent systems, data-driven modeling, and fault diagnosis.



FERNANDO SANCHE received the Ph.D. degree in mathematics from the Universidad de Sevilla, Seville, in 2002, where he is currently a Professor with the Department of Computer Science and Artificial Intelligence. His areas of interest include the mathematical study of complex networks as universal representation systems for highly complex systems, and the mathematical analysis and control of multi-agent systems formed by especially simple units. He also makes use of machine learning techniques for modeling, prediction, and the control of complex dynamical systems, with a particular emphasis on those that have a very high-semantic load that makes it difficult to process them by means of the most common data analysis techniques.



MARIELA CERRADA received the B.Sc. degree in system engineering and the M.Sc. degree in control engineering from the University of Los Andes (ULA), Venezuela, in 1993 and 1998, respectively, and the Ph.D. degree from the National Institute of Applied Sciences, Toulouse, France, in 2003. She was with the Control Systems Department, ULA, from 1994 to 2016. She was a Prometeo Researcher with Universidad Politécnica Salesiana (UPS), Cuenca, Ecuador, from 2014 to 2015 and from 2016 to 2017. She has been with the Mechatronics Department, the UPS, since 2017. Her main research interests include industrial intelligent supervision, fault diagnosis, prognostics, and health management.



CHUAN LI (M'16–SM'18) received the Ph.D. degree from Chongqing University, China, in 2007. He has been a Postdoctoral Fellow with the University of Ottawa, Canada; a Research Professor with Korea University, South Korea; a Senior Research Associate with the City University of Hong Kong, China; and a Prometeo Researcher with the Universidad Politécnica Salesiana, Ecuador. He is currently a Professor with the School of Mechanical Engineering, Dongguan University of Technology. His research interests include prognostics and health management, industrial engineering, and intelligent systems.

...